DOCUMENT RESUME

ED 247 235 TM 840 363

AUTHOR	Haenn, Joseph F.; And Others
TITLE	Effective Score Reporting of Non-Norm-Referenced Assessment.
INSTITUTION	Measurement Inc., Durham, NC.; RMC Research Corp., Arlington, Va.
SPONS AGENCY	Maryland State Dept. of Education, Baltimore.
PUB DATE	Apr 84
NOTE	33p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984). Small print in Appendix A.
PUB TYPE	Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160) -- Information Analyses (070)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Criterion Referenced Tests; *Information Utilization; *Minimum Competency Testing; *Scores; Secondary Education; Surveys; Test Construction; Testing Programs; Test Results; *Test Use
IDENTIFIERS	Maryland Functional Testing Program; *Test Reporting

ABSTRACT

The Maryland State Department of Education issued a request for proposals to develop a score reporting system for the Maryland Functional Testing Program. RMC Research Corporation conducted a literature review of extant literature and developed a national survey of non-norm-referenced test score reporting practices. This comprehensive analysis of current practice in, and research on, the reporting of results of functional testing programs was conducted under contract to Measurement, Incorporated. A literature review revealed limited material about score reports for functional testing programs. The literature recommended that score reports include a testing program description, presentation of results in an audience-appropriate manner, and inclusion of cautions about the limitations of interpretation inherent in the score. A national survey of state and local score reporting practices was designed. On the basis of survey results, several recommendations were made concerning simple, concise, and clearly labelled report design; inclusion of only that information necessary to the user; support documents; interagency communication; standardization of test results interpretation; narrative results summaries; and classification data for district comparisons. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED247235

Effective Score Reporting of Non-Norm-Referenced Assessment

Joseph F. Haenn
North Carolina Office of The Governor

Michael B. Bunch
Measurement Incorporated

C. Wain Mengel
RMC Research Corporation

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. F. Haenn

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 840 363

Paper presented at the Annual Meetings of the
American Educational Research Association
New Orleans, April, 1984.

Executive Summary

In March 1983, the Maryland State Department of Education (MSDE) issued a request for proposals to develop a score reporting system for the Maryland Functional Testing Program. A contract was subsequently awarded to Measurement Incorporated (MI) of Durham, North Carolina. Central to the execution of this contract was comprehensive analysis of current practice in and research on the reporting of results of functional testing programs. RMC Research Corporation (RMC) was subcontracted to conduct a thorough review of extant literature on the subject and to conduct a national survey of non-norm-referenced test score reporting practices.

Literature Review

Very little has actually ever been published about score reports for functional testing programs. Most of the literature deals with the types of scores that can or should be reported and how those scores should be interpreted. Few publications address how to make results meaningful and useful to users. However, based on the available literature, and from reporting guidelines applicable to all types of tests, three specific recommendations can be made:

1. Every score report should contain a description of the testing program. Specific details of the program will vary from one score report recipient to another.

2. Scores should be presented in an audience-appropriate manner. In other words, since students and superintendents have different information needs, they should get different reports.

3. Each score report should contain cautions about the limitations of interpretation inherent in the score.

The first section of this paper provides detailed explanations of each of these three recommendations.

National Survey

The information gathered in the literature search guided the development of a 34-item questionnaire adapted for use at the state and local levels. Directors of testing were asked questions about five

major topics: 1) total test scores, 2) subtest scores, 3) subject matter summary scores, 4) specialized services, and 5) standard services. Directors were asked to indicate whether various score reporting features were included in reports sent to pupils, teachers, principals, district administrators, and state agencies. Questionnaire responses were obtained from 28 states and 27 districts.

Project staff conducted telephone interviews with approximately 20 state and local directors of testing to determine specific aspects of test score reporting practices. These interviews also helped to increase response rate. Finally, all respondents were asked to send sample score reports and support documents. Samples were received from 14 states and 10 local districts.

Survey results showed that the state-of-the-art in functional test score reporting is primitive. While several (about 23-30) features can be identified as potentially appropriate to one audience or another (e.g. pupils, principals, etc.), the typical score report contains only about one-fourth of them. In general, the further removed the report is from legislated or mandated state reports, the less information the report has and the less interpretation is provided. Thus, the typical student report contains only about 18 percent of the information it might, while the typical state report contains approximately 32 percent of the information it could conceivably contain. Specificly noteworthy shortcomings include lack of narrative reports and lact of reporting on previous performance (to provide an historical perspective for interpretation).

While reports and reporting systems in general had serious shortcomings, several specific reports and support documents provided much information in clear and meaningful ways. However, a local survey should be used to determine local user information needs.

Eight specific recommendations were developed based on the national survey. These are summarized below.

1. Reports should be simple, clear, and clearly labelled as to intended audience.
2. Each report should contain no information not specifically intended for the specified audience.
3. Support documents should receive just as much attention in development and use as do score reports.
4. Communications with other state agencies and media about test scores and the testing program are crucial.
5. Dissemination and interpretation of results should be standardized through workshops and other means.
6. Past performance should be reported at least at the district and school levels.
7. Narrative summaries should be provided at all reporting levels.
8. Inappropriate district comparisons should be discouraged and appropriate comparisons facilitated by providing information about district comparability.

TABLE OF CONTENTS

	Page
EXECUTIVE SUMMARY	1
Literature Review	1
National Survey	1
PART I: LITERATURE REVIEW	1
Overview of Score Reporting	2
General Considerations in Score Reporting	2
Specific Recommendations in Score Reporting	3
PART II: NATIONAL SURVEY	6
Design	6
Results	8
PART III: CONCLUSIONS	15
REFERENCES	17
APPENDIX A: NATIONAL SURVEY QUESTIONNAIRES	19
State Score Reporting Survey	20
School District Score Reporting Survey	23
APPENDIX B: FOLLOW-UP TELEPHONE INTERVIEW FORMS	26
Follow-up Questions/State	27
Follow-up Questions/District	28

PART I: LITERATURE REVIEW

The Request for Proposals (RFP) for "The Development of the Score Reporting System for the Maryland Functional Testing Program" called for a report on the state of the art in score reporting especially for non-norm-referenced testing programs, including information from both available research literature and a national survey of the score reporting systems used across the country. This part of the report provides preliminary information on score reporting as documented in journal articles, ERIC files, and from a limited number of fugitive documents (e. g. papers, research bulletins, and other unpublished materials). It was expanded through the national survey of score reporting systems and the collection of documents obtained through that survey.

This literature review is brief for several reasons. First, textbooks were not included in the review because most provided generic answers to score reporting situations, with few applications for non-norm-referenced testing situations. Second, only a few fugitive documents were available and included in this preliminary literature review. In any state of the art review, fugitive documents provide the most recently available literature because of the time delay of submitting manuscripts or articles for journals. Finally, few articles were available through the Educational Resource Information Center (ERIC). This is due to two factors. First, only recent articles (since January, 1980) were included in the review. Also, there is a lack of adequate descriptors for this topical area, due in large part to the absence of a need for such descriptors because of limited availability of documents on this topic. The ERIC descriptors used were:

Criterion Referenced Tests
Information Utilization
Norm Referenced Tests
Scores
Test Results

The ERIC identifiers which appear to be most relevant are:

Test Scoring
Test Reporting

The rest of this part of the report briefly reviews the available literature about the state of the art in score reporting. The first section overviews some of the basic problems in score reporting. Following are sections on general considerations in score reporting and specific recommendations for score reporting.

Overview of Score Reporting

The lack of a great deal of information about score reporting is due to the only recent emergence of test score reporting to the public, especially about criterion-referenced testing. As recently as ten years ago, only about half of the school districts in the United States with enrollments of more that 12,000 students released test score results (ERS, 1974). Then, as cries for accountability in education led to increased score reporting, the National School Public Relations Association released a document explaining how test scores should be released to the public (NSPRA, 1976).

General Considerations in Score Reporting

Unfortunately, much of the reporting which does occur involves the straightforward reporting of scores, with little attention paid to score interpretation or ways of using the results for instructional or program improvement (Lyon et al, 1978). The overall use of test scores is dependent upon adequate test score reporting.

The purpose of a testing program is, after all, to provide useable information in a convenient format. Test score information that is inappropriate, confusing, or in any other way unsuited to the needs of potential test score users will be of limited value (Mills and Hambleton, 1979, p.1).

Numerous authors have reported on the misuse of test scores (Hoffman, 1962; Wargo and Green, 1978; Fisher, 1978; NEA, 1979). However, at least some of this misuse is due to the manner in which scores are reported. The clarity of reporting (Alkin et al, 1979) and the feasibility of using the information which is reported are two factors which determine whether or not or to what degree test results can (will) be used.

Interpretations of test score results also can increase their utility to a great degree (Fiechtling and Myerberg, 1983). Different reports should be generated for different audiences, considering the needs of each audience and ensuring that adequate information is provided to allow for proper score result interpretation. The computer has facilitated greatly the task of scoring and generating targeted score reports (Mills and Hambleton, 1980).

A single test score can be interpreted in a number of different ways, both for norm-referenced and criterion-referenced purposes (Mills and Hambleton, 1979). The criterion-referenced purposes include diagnosis (what a student can and cannot do), research (progress of

experimental and non-experimental groups), evaluation (measurement of progress relative to set of objectives), and assessment (achievement of a set of competencies).

Specific Recommendations in Score Reporting

When reporting test scores to any audience, there are three points which should be included: a description of the testing program, presentation of the test results to specific audiences, and cautions to be observed in interpreting these results (Frechtling and Myerberg, 1983). These three points need to be emphasized to different degrees depending upon the specific situation. They form the focus for the remainder of this discussion.

Description of the testing program. Before presenting any test results, the context of the testing program should be established. The reason(s) for the testing program, including the uses of the testing results, should be presented. The name(s) of the test(s) used, including the form(s) and level(s), subtests given, publication or standardization date(s), and administration date(s) should be indicated as well as other testing information such as grade levels tested, types of students tested and exempted from the testing program, and any testing situation changes as compared to previous administrations. The type of content tested should be explained in sufficient detail for the intended audience including skills tested, item formats and number of items for each skill or other score level. The types of scores reported should be described including the name of the score metric, the type of interpretation (e.g., norm-referenced) and the advantages and limitations of the score metric. In other words, the tesing program should be explained in detail appropriate to the sophistication of the intended audience.

Presentation of test results. The presentation of test results can occur in different forms depending on the reporting audience and the level of aggregation. There are five primary levels of aggregation: student level, classroom level, school level, district level, and state level. There are numerous reporting audiences, but the primary ones are parent (or student or counselor), teacher, principal, superintendent or board of education, and research and evaluation staff. Few test reports other than data tabulations and displays are prepared for research and evaluation personnel. The other four audiences primarily correspond to the four levels of aggregation. However, most audiences also can use reports relevant to a lower level of aggregation (e. g., teachers can use individual student reports; principals can use classroom reports; superintendents can use building reports.)

Student level results usually are presented to parents. The description of the testing program should be brief and easy to read. Forms which have been successful include questions and answers and brochures. Individual student results can be incorporated in the overview document or presented separately. The choice of which metric

to report is very important, as some metrics need much more interpretive information to be used correctly (e.g., grade equivalents). Since a test score can represent a multiplicity of factors, as much background information should be provided as is practical (Performance Printouts for Parents, 1974). Also, scores should be represented as score bands instead of discrete points so as not to mislead interpretation as a finite measure (Backman, 1976; Frechtling and Myerberg, 1983). Where possible, individual student results should be compared with others (e.g., class average, grade average), with the objectives requiring additional study indicated along with suggested study materials (Furlong and Miller, 1978). Anastasi (1976) strongly recommends reporting broad levels of performance and qualitative descriptions of results, fully interpreted in words, rather than the presentation of numerical scores. This idea of highlighting student strengths and weaknesses in text has been supported by others (Frechtling and Myerberg, 1983). The emphasis should be on communication, not jargon. These reports also can be used by guidance counselors.

Classroom level results usually are presented to teachers. Teachers need this information for decison making, but often have a limited background and familiarity with the field of testing and measurement (Goslin, 1967). The information needs of teachers may be dictated by school district conventions or current concerns (emphasis on spelling or vocabulary). One type of report prepared for teachers is a preformance-by-objective report focusing on specific items (Allen and Fredrick, 1979), including the number of items for each objective, item difficulties, and relevance of items to objectives emphaized by the school or district. These results can be used to guide classroom teaching including student grouping and content emphases.

School level reports are used primarily by principals. These reports can be used to determine school-level strengths and weaknesses and, when examined longitudinally, can assess progress or change. Such information should receive limited distribution as it often is misintrepreted. Ability and/or socioeconomic status differences among schools have a large impact on school level data. Lists of high and low performing students can be used for student selection for special programs. Other types of reports at the school level include performance by feeder school, performance across years, or performance by special program.

District level reports are used primarily by specialists plus the district superintendent and the board of education. Sophistication of the intended audience will affect the level of detail of these reports. Results may be broken down by sex, ethnicity, socioeconomic status or other sensitive level of aggregation, but such breakdowns should include considerable cautions in interpretation. A report to the school board should be more general and not focus on specific items (Allen and Fredrick, 1979). Average student performance, variance among students, comparison of student performance to that of previous years, and other historical references can be included in these reports.

State level results are primarily intended for use by legislators, although there are the inevitable district-by-district comparisons.

More important is the comparison of each district to similar districts (e.g., similar socioeconomic status or size) within the state.

Interpretation cautions. Several examples of cautions to be observed in interpreting data at the various levels of aggregation were explained in the preceeding section. It must be emphasized that misinterpretation of test results by well meaning individuals leads to the greatest misuse of test score results. One can never assume that educators understand all of the ramifications of the various test score metrics and graphical displays. All results should be fully interpreted within the text of the report.

PART II: NATIONAL SURVEY

In the development of a new score reporting system, it is wise to assess the current state of the art in score reporting. The request for proposals (RFP) for this project indicated two ways to determine the state of the art: a literature review and a national survey of state and local education agencies. This part of the report summarizes the results of a national survey of state and local score reporting practices and describes the procedures by which that survey was undertaken. Results of this report were intended to guide development of the reporting system for the Maryland Functional Testing Program.

This part of the report describes the survey design and results. The design section describes in some detail the procedures used in collecting information about state and local score reporting practices. The results section contains a summary of the information collected.

Design

The survey was designed around a questionnaire, adapted for state and district audiences, and mailed to a sample of 49 state directors of assessment and 57 district directors of assessment, testing, evaluation, or equivalent position. Many respondents also participated in follow-up in-depth interviews.

Sample. Questionnaires were mailed to 49 states (Maryland was excluded) and to 57 district testing directors. The reason for sending so many district questionnaires was to compensate for an anticipated poor response rate. We expected a 50-60 percent response rate because the questionnaires were sent at a time when most local school districts would be starting the school year and would be under tremendous pressure to deal with issues such as physical plants, bus schedules, student transfers, and other matters that always seem to require extra effort in the first few weeks of a new school year. This oversampling was designed to yield at least 20 district responses. States, of course, could not be oversampled since our sample was actually the entire universe of states.

The sample of states and districts for follow-up interviews was selected on the basis of whether or not the state had a non-norm-referenced test. Thus, in states where no statewide test was in place, neither the state director of assessment nor any local director was

interviewed. From the list of states with statewide testing programs 15 states and 15 districts were selected in such a way that all regions of the country would be represented and so that the sample would contain both large and small states and states and districts which appeared to have testing programs that merited a closer look.

Questionnaire design. A list of questions to guide the development of criterion-referenced test score reports and a set of questions about test score reports were developed based on a paper prepared by Craig Mills and Ronald Hambleton (1980). The questions in that paper actually constituted a checklist or rating form for score reporting systems. Two prototype questionnaires (one for states and one for districts) were developed based on this rating form. From 69 items in the Mills and Hambleton form, 34 items were identified that addressed features the Maryland State Department of Education staff considered important. These two questionnaires are included as Appendix A.

While each questionnaire is included in its entirety in Appendix A, it seems worthwhile to provide a summary description here. Each version of the questionnaire consists of a cover sheet and a two-page list of specific questions. The cover sheet contains basic questions such as whether or not the state or district has a criterion-referenced testing program (Section A), which grades are tested, and what types (levels) of reports are generated (Section B). The two-page questionnaire (one legal-size sheet, front and back) is divided into five sections. These remaining sections may be summarized as follows:

Section C - Total test
Section D - Subtests or objective clusters
Section E - Subject matter summary reports
Section F - Specialized services
Section G - Standard services

The sections range in length from 4 to 13 questions. Questions are essentially the same for state and for district questionnaires but are worded a little differently to make them audience appropriate.

Respondents were asked to respond in one of two ways. If they had no testing program, they were asked to answer the questions in terms of what should occur. If they had a testing program, they were asked to respond in terms of actual practice. State directors of assessment were asked to answer each question as it related to individual pupils, classes, buildings, districts, and state reports. District directors were asked to answer each question as it related to pupils, classes, buildings, and the district. All respondents were asked to send copies of score reports, support materials or other items they considered appropriate.

Follow-up interviews. As noted previously, 15 states and 15 districts were selected for follow-up interviews with the directors of assessment. Questions were based on information obtained from the questionnaire as well as the general recommendations included in the literature review. A core of four major questions provided the framework for each interview. For district directors of assessment,

these basic questions dealt with use of score reports, types of questions asked by parents and others, relevance of test scores to local curriculum, and types of misinterpretation occurring in the district. The questions for state directors of assessment were similar and included questions about the type of interpretation support offered and recent changes, if any, in the score reporting system. Copies of the interview forms are in Appendix B. Each interview began with these basic questions and then took whatever direction seemed appropriate, given respondents' comments.

Data analysis. Questionnaire responses were coded and tabulated to yield response rate and frequency distributions for each item. The responses of individuals in states or districts which have testing programs were compared to those of individuals in states or districts with no testing programs. Analysis of telephone intereview data was restricted to counting common comments, complaints, and concerns. In all of the analyses, we attempted to note major trends, as well as to detect interesting and useful practices unique to individual states or districts. No elaborate cross-tabulation of questionnaire responses was attempted since straightforward reporting of the results was both appropriate and sufficient to guide development of instruments to be used in a subsequent survey of Maryland audiences and eventually in the design of score reports.

Results

Response rate. We received completed questionnaires from 28 states and 27 districts. The state questionnaire response rate (57 percent) is consistent with the predicted 50-60 percent response rate. Of the 21 nonresponding states, two actually returned the first page of the questionnaire indicating that they had no test. Follow-up telephone calls to the remaining 19 states revealed that eight had no statewide testing program and eleven had testing programs. Ultimately, only five states with testing programs failed to produce a response.

With respect to district questionnaire responses, it seems that oversampling had the desired effect. Follow-up telephone calls to nonresponding districts revealed that, indeed, the timing of the survey was less than optimal. The most frequently cited reason for not having responded was either lack of time or lack of staff. The 27 district responses are, however, amply augmented with interview data.

State responses. Table 1 on the following page is a tabulation of questionnaire responses of state directors of assessment. The table is divided into two sections, the left section representing states with criterion referenced assessment programs, and the right section representing those without such programs. Two facts emerge when the two groups are compared. First, assessment directors in states presently with no programs indicated they would report approximately twice as much data as is being reported in states with programs. This finding is not suprising since the questionnaire encouraged respondents to list what should be in reports rather than what is in them.

Table 1
Summary of Questionnaire Responses by State Directors of Assessment
(Percentage of Directors Choosing Each Response)

Statewide Testing Program in Force (Responses indicate actual practice) N=21						No Statewide Testing Program (Responses indicate what should appear on hypothetical report) N=7				
Level of Report						Level of Report				
Pupil	Class	School	District	State	Item*	Pupil	Class	School	District	State
38	33	52	43	52	C1	43	43	71	100	86
19	24	33	24	24	C2	29	43	86	57	57
24	29	29	29	29	C3	71	71	43	14	14
19	33	24	19	5	C4	29	86	29	0	0
24	24	24	14	14	C5	57	43	0	0	0
14	33	62	48	52	C6	14	43	57	43	43
48	38	43	38	33	C7	57	29	29	14	0
29	33	48	48	48	C8	29	43	71	57	57
14	29	43	43	38	C9	29	43	71	86	43
10	38	67	57	57	C10	14	43	86	100	86
10	5	14	14	19	C11	14	29	57	71	57
0	10	24	29	24	C12	0	14	43	43	29
10	14	24	19	14	C13	0	14	57	71	29
43	24	19	14	24	D1	71	100	29	14	14
10	24	43	43	43	D2	29	57	57	43	29
19	24	33	29	29	D3	43	57	43	14	0
0	5	14	14	19	D4	14	29	43	29	14
10	14	10	14	19	D5	14	43	71	71	29
5	5	14	10	10	D6	0	14	43	57	29
43	33	24	19	10	E1	100	71	43	0	0
5	19	48	52	62	E2	29	57	57	43	14
5	5	24	24	33	E3	14	43	43	14	14
5	5	14	14	14	E4	14	14	43	57	14
10	5	5	10	24	F1	14	0	14	43	43
10	5	0	5	10	F2	43	0	0	0	0
10	5	0	5	14	F3	0	14	43	29	29
48	14	5	0	0	F4	57	43	29	14	0
5	5	5	5	5	F5	0	29	29	14	0
33	33	48	48	57	G1	57	71	86	86	57
5	10	19	19	24	G2	57	29	43	57	14
43	33	29	29	29	G3	29	29	43	57	43
10	10	10	24	24	G4	14	14	43	29	29
14	14	19	24	43	G5	43	29	71	57	43
24	10	19	29	48	G6	57	57	71	43	43

* See Appendix A for a complete description of each item.

The second finding was surprising. At every level except one, the "no-program" assessment directors checked twice as many options as the "program" assessment directors. The exception was the state level. The responses were essentially the same when one compares only at the state level. The directors of assessment in states with no criterion-referenced testing programs would report far more information to their districts, buildings, classrooms, and pupils as do their counterparts in states with intact criterion-referenced testing programs. Yet they would not increase the information load at the state level.

Turning our attention now to what is actually reported at the state level, we find that states provide the most information at the state level, slightly less at the district and building level, and much less at the classroom and pupil level. Looking only at the items labelled "possible information need" by Mills and Hambleton, we derive the information "scores" depicted in Table 2.

Table 2

Mean Raw Scores of Statewide CRT Reporting Systems
By Type of Report*

Type of Report	Possible Score	Mean Score	Percent of Possible
Pupil	26	4.7	18.1
Class	30	5.5	18.3
School	26	6.4	24.6
District	23	6.4	27.8
State	23	7.5	32.6
Total	128	30.5	23.8

The scores shown in Table 2 show that the typical state criterion-referenced testing reporting system has a long way to go in meeting possible information needs of local and state constituents. Even in states where there is no statewide criterion-referenced testing, state directors of testing, who were encouraged to respond in ideal terms, could only envision systems which could meet about 40 percent of state information needs, 57 percent of district needs, 54 percent of school needs, 41 percent of classroom needs, and 33 percent of pupil and parent information needs.

* Based on items considered "possible information need" (Mills and Hambleton, 1980):
  Pupil (all items except C3, C4, C8, C10, C12, D4, D5, F4)
  Class (all items except C7, C8, F2, F4)
  Building (all items except C3, C4, C5, C7, D1, D3, E1, F2, F5)
  District (all items excpet C3, C4, C5, C7, D1, D2, D3, E1, F2, F4, F5)
  State (all items except C3, C4, C5, C7, D1, D2, D3, E1, F2, F4, F5)

There were no universal practices. Every feature generated a variety of responses. The most commonly reported features were as follows:

Number of items tested per objective
Number or percent of masters by objective
Average scores for each subject
Audience identification (Building, District, and State levels only)

Conversely, several features were chosen by only a few respondents (25% or fewer):

Diagnostic statements about performance on non-mastered objectives
Past performances on the same or related objectives
Normative information at any level
Past performances on objective clusters
Objective clusters on which overall performance was low
Statements from state officials
Teacher comments
Information on "key objectives"
Performance relative to aptitude
Generalizability of scores

District responses. If one considers the information listed above to be important and necessary at some level of score reporting, and if state reports lack most of it, then local school districts must fill in the information gaps. A summary of the school district questionnaires returned shows that the average school district provides precisely the same amount of district level information as the average state, very slightly more information at the building level, and about half again as much at the classroom and pupil level as does the "average" state. Again, however, there are some large reporting gaps. For example, fewer than one fourth of the assessment directors reporting an intact criterion-referenced testing program routinely supply schools, classrooms, or pupils (or their parents) with the following information:

Diagnostic statements about performance
Past performances on the same or related objectives
Normative information at any level
Past performance on objective clusters
Objective clusters on which overall performance was low
Statements from district officials
Teacher comments
Information on key objectives
Performance relative to aptitude
Generalizability of scores

This list is, of course, the same as the one for state reports shown above. In fact, there were additional gaps in district level reporting that indicate that districts report even less to classrooms and students than do states. Other items of information reported by fewer than 25% of reporting districts were as follows:

Responses chosen for incorrectly answered items
Average scores by objective
Frequency distribution of percent of objectives mastered
Almost any information on clusters of objectives
Audience identification
Phone number for interpretation help

Very few of the features listed in the district questionnaire were checked by 50 percent or more of respondents. Those items indicated by half or more of district directors of assessment as appearing on school, classroom, or pupil score reports included the following:

Names of masters and non-masters (classroom level only)
Mastered and non-mastered objectives
Number or percent of masters by objective (classroom and school levels)
Number of students tested on each objective (classroom and school levels)
Individual student scores for each subject

Telephone interviews. The telephone interviews yielded perhaps the most useful and interesting information of all. State and local directors of assessment revealed a variety of unique practices and applications for their testing programs. Table 3 summarizes some of the uses of results, questions about results, and misuses or misinterpretations of results.

Table 3
Summary of Telephone Interview Data

Use of Results	
by States:	State reports (mandates by legislation)
	Reports to districts (data only)
	Guidance for districts
	Planning and budgeting
	Curriculum review
	Other political purposes
	Data source for research grants
by Districts:	Textbook/curriculum evaluation
	Establishing instructional priorities
	Student profile generation
	Comparison of schools
	Generation of various lists
by Schools:	Teacher/grade reports
	Class profiles
	Identification of individual student strengths and weaknesses
	Scheduling for special courses
	Goal and priority setting

Table 3 (Cont.)
Summary of Telephone Interview Data

Questions and Results	
from Parents:	How can one test keep my child from graduating/being promoted? Is this test required for graduation? How can a person get X% and still fail? How are the tests scored?
from Teachers:	How do I use this?
Misuses/Misinterpretations/Abuses	Problems in using and converting various metrics (e.g. raw score, scale score, etc.) Equating minimum competence with fulfillment of all goals of schooling Class, school, and district comparisions without regard to resources Cheating

Several interesting practices were also revealed. Listed below are some of the major findings.

- Some states use a National Assessment of Educational Progress approach wherein only one or two subjects are tested each year. Over a 3-5 year cycle, all subjects are tested. One variation has Reading and Mathematics tested each year with one other subject.

- Some states provide optional services to local school districts at district expense. These services include data tapes, detailed profiles, extra levels of reports.

- Some states have developed excellent rapport with the press and other media and routinely report to them with the same attention to detail and audience appropriateness as they give other users.

- One state requires local districts to print and score the tests at local expense and to report to the state. This policy will probably change in the next year.

Sample documents. Fourteen states and ten districts provided smaples of score reports, user's guides, or other documents. Reports range from black-on-white, computer-printed documents to preprinted, color-coded reports. Several student reports contain a narrative summary. These tend to be fairly straightforward and offer little or no interpretation. One student report is written in both English and Spanish. Stick-on labels are provided by somes states. Georgia, for example, has a pressure-sensitive label listing scores in Reading and Mathematics and in several subtests. This label is presumably for a permanent record folder.

The most appealing features of student level reports may be summarized as follows:

- Easy to read
- Description of testing program
- Some diagnostic information
- Student performance summarized in words as well as in numbers

Georgia, Missouri, North Carolina, South Carolina and West Virginia plus the school districts of Fairfax County, VA, New York City, and Palm Beach County, FL, provided sample reports.

Desirable characteristics of class reports may be summarized as follows:

- Easy to read
- Indication of class strengths and weaknesses
- Domains described or defined
- Caveats regarding interpretation

Missouri, South Carolina and West Virginia plus the school districts of Dallas, Texas, Fairfax County, New York City and Palm Beach County provided sample reports.

Desirable features of school and district reports include the following:

- Test overview
- Description of scores
- Interpretation
- Scores represented as ranges, not points
- Subgroup analysis
- Directions for further analysis

California, Missouri and South Carolina plus the school districts of Fairfax County, New York City and Palm Beach County provided sample reports.

Interpretation Guides were sent by directors of assessment in nine states and two districts. One of the district manuals is very helpful and concise. It is not elaborate or sophisticated. It conveys basic information in tabular form, provides a program overview, and describes the results in narrative form. The best example of a statewide interpretation guide comes from Texas. This publication is sophisticated but very conspicuously oriented toward a user audience. It not only contains essential information, but it presents the information in a highly readable and inviting fashion.

PART III: CONCLUSIONS

In this final part of the report, we draw several conclusions regarding information needs, report design, report content, support documents and services, and communications and public relations. We recommend general reporting system features rather than specific ones. More specific recommendations should be based on the needs of local score report users.

The current state of the art in criterion-referenced test score reporting is far from what it could be. The typical state and district report only a small portion of what educators need in order to use test results effectively. Even the most ambitious states and districts report only about half of what current research indictes could be reported. One major concern is an explanation of test scores (cf. questionnaire item G6). Ten out of 21 states provided such narratives in state reports but only five provided such information to students, and only two provided narrative summaries of data to teachers. Another concern is information on past performances (C11, D4, E3). Virtually no state reports information about past performances at any level. Although seven states (out of 20) report past performances on domains or subject matter areas at the state level, only one does so at the classroom level, and only five do so at the school and district levels.

There are, however, bright spots which stand out very clearly against this dark background. Some states are pursuing a vigorous campaign of educating test users through audience-appropriate publications and workshops. A few states have scaled back the overall testing program to a cyclical review process and are now able to concentrate more on use and interpretation. Many states and some districts provide good score reports at pupil, class, school, and district levels, even if these reports do not contain all the characteristics of an ideal report.

It should be clear that each state with a functional testing program has different information needs at all reporting levels. It would be unwise indeed simply to take the best features of several reports and mold them into a reporting system. Instead, we urge the states and local districts to consider some of these exemplary practices and products in light of their own unique needs.

Based on the results of the national survey, the following general recommendations can be made:

- At all levels, the overall report design should be simple, concise, and clearly labelled as to its intended audience and purpose.

- Each report should contain only that information which is necessary for the user (student, parent, teacher, etc.) to make informed decisions, but should contain sufficient information and caveats to guide those decisions. Specific suggestions by audience have been mentioned in the body of this report.

- Support documents (user's guides, technical manuals, brochures, etc.) are just as crucial as the reports themselves, if not more so.

- Communication with other agencies and the media is also crucial. Statewide acceptance and understanding are essential to appropriate interpretation and use of test results. Examples are available.

- Because persons other than state staff are often involved in disseminating results, attempts should be made to standardize the interpretation of results through workshops and other means. In order to facilitate the charting of progress at many levels, reporting past performances should be at least considered at the district and school levels.

- Narrative summaries of results should be provided at every reporting level.

- Because district comparisons are inevitable, classification data (e.g. size, per pupil expenditure (etc.) should be provided which would at least suggest which comparisons are appropriate and which are clearly not.

As the reader can see, some of these recommendations go beyond the actual score report and its support guide and manuals. The reason for this is quite simple. Score reports are interpreted in a specific context which varies from district to district, from school to school, from teacher to teacher, and from student to students. Fostering a general concern an appreciation for accuracte, reliable, and meaningful score reporting should be a task to be executed in every aspect of any testing program.

References

Alkin, M., Duillak, R., and White, P. Using Evaluations: Does Evaluation Make a Difference? Beverly Hills, CA: Sage Publications, 1979.

Allen, R. and Fredrick, V. Testing Services Manual. Wisconsin State Department of Public Instruction. Madison, WI, 1979. (ED 190 667)

Anastasia, A. Psychological Testing (4th edition). New York: MacMillan, 1976.

Backman, M. "Reporting test results for guidance and instructional purposes." In Guide for School Testing Programs. Washington, D.C.: National Council on Measurement in Education, 1976.

Educational Research Service (ERS). Releasing Standardized Test Scores to the Public, 1974.

Fisher, T. H. "The assessment of student competencies in Florida's public schools." In R. B. Ingle, M. R., Carroll, and W. J. Gephart (Eds.), The Assessment of Student Competence in the Public Schools. Bloomington, IN: Phi Delta Kappa, 1978, pp. 24-38.

Frechtling, J. A. and Myerberg, N.J. Reporting Test Scores to Different Audiences. ERIC/TM Report 85. Princeton, N.J., December 1983.

Furlong, F. and Miller, W. "DIAGNOSE: Computer-based reporting of criterion-referenced test results." Educational Technology. 1978, 18: 37-39.

Goslin, D. Teachers and Testing. New York, Russell Sage Foundation, 1967.

Hoffman, B. The Tyranny of Testing. New York: Crowell-Collin-MacMillan, 1962.

Lyon, C., Doscher, L., and McGranahan, P. Evaluation and School Districts. Center for the Study of Evaluation. Los Angeles, CSE/UCLA Graduate School of Education, December 1978.

Mills, C. N. and Hambleton, R. K. "Guidelines for reporting criterion-referenced test score information." Paper presented at the Annual Meeting of the American Educational Research Association. Boston, MA. April 1980. (ED 189 130)

National Education Association (NEA). Testimony by Executive Director Terry Herndon before the Senate Subcommittee on Education. Washington, D.C.: Author, June 19, 1979.

National School Public Relations Association (NSPRA). Releasing Test Scores: Educational Assessment Program, How to Tell the Public. Arlington, VA: Author, 1976.

"Performance printouts for parents." Nation's School and Colleges, September 1974, 1:31-32.

Wargo, M. and Green, D. Achievement Testing of Disadvantaged and Minority Students for Educational Program Evaluation. New York: McGraw-Hill, 1978.

APPENDIX A

NATIONAL SURVEY QUESTIONNAIRES

STATE SCORE REPORTING SURVEY

Name of Respondent (Optional): _____

Title/Responsibilities: _____

Does your state have a non norm-referenced (e.g., domain-referenced, objective-referenced, or criterion-referenced) district-wide assessment program?

Check one Box: [A1]

☐ Yes

☐ No

DIRECTIONS:

Please take a few minutes to answer the following survey. If you answered "no" to the question above, please skip to page 2 of the survey and respond according to what you, personally and professionally, believe would make for a good score reporting system for your state. If you answered "yes" to the question above, please complete the section below before completing the remainder of the survey. Complete the survey based on the score reporting system currently being used by your state.

Complete the following questions on this page only if you answered "yes" to the question above.

Circle the grade levels assessed (circle as many as apply):

PreK K 1 2 3 4 5 6 7 8 9 10 11 12 [B1-14]

Check the types of reports generated (check as many as apply):

☐ Classroom teacher reports

☐ Individual student (parent reports)

☐ Individual subject matter summary reports

☐ Building level reports

☐ District level reports [B15-20]

☐ State level reports

Please add any comments below regarding your score reporting system which might be useful in interpreting your responses to this survey.

DIRECTIONS

If you answered "yes":

For the rest of the survey, respond to each item in terms of what your reporting system does. Respond for each level of reporting (e.g., teacher report, building report) by placing a check mark ("✓") in the appropriate boxes if this item is part of your score reporting system.

If you answered "no":

Since you do not have a score reporting system, you should place a check mark ("✓") in the appropriate box(es) for each level of reporting (e.g., teacher report, building report) if you feel that item would be desirable in a score reporting system.

Please enclose a copy of any score reporting forms you are currently using. THANK YOU for your participation in this effort. Please return this form and any score reports you choose to send to:

Dr. Joseph F. Haenn
Senior Research Associate
Maryland Score Report Development Project
RMC Research Corporation
108 S. Churton Street
Hillsborough, NC 27278
(919) 732-6121

Pupil (Parent) Report	Class (Teacher) Report	Building Report	District Report	State Report	Items: In order to meet our informational needs, does (should) the score reporting system at each indicated reporting level. . .	
P	C	B	D	S		
☐	☐	☐	☐	☐	indicate the number of items tested per objective?	[C1]
☐	☐	☐	☐	☐	indicate the cut-off score for each objective?	[C2]
☐	☐	☐	☐	☐	identify the responses chosen for incorrectly answered items?	[C3]
☐	☐	☐	☐	☐	report the names of masters and non-masters by objectives?	[C4]
☐	☐	☐	☐	☐	provide diagnostic statements about performance on non-mastered objectives?	[C5]
☐	☐	☐	☐	☐	report number or percent of masters for each objective?	[C6]
☐	☐	☐	☐	☐	identify mastered and non mastered objectives?	[C7]
☐	☐	☐	☐	☐	report objectives on which performance was low?	[C8]
☐	☐	☐	☐	☐	report average for each objective?	[C9]
☐	☐	☐	☐	☐	report the number of students tested for each objective?	[C10]
☐	☐	☐	☐	☐	report past performances on the same or related objective?	[C11]
☐	☐	☐	☐	☐	report frequency distribution of percent of objectives mastered?	[C12]
☐	☐	☐	☐	☐	report normative information for each objective?	[C13]

Please turn the page.

Pupil (Parent) Report P	Class (Teacher) Report C	Building Report B	District Report D	State Report S	
					The following items refer to the reporting of clusters of objectives or subtests. Does (should) the report system. . .
☐	☐	☐	☐	☐	report individual scores on each objective cluster? [D1]
☐	☐	☐	☐	☐	report average score on each objective cluster? [D2]
☐	☐	☐	☐	☐	report number of percent of objectives mastered for each objective cluster? [D3]
☐	☐	☐	☐	☐	report past performance on each objective cluster? [D4]
☐	☐	☐	☐	☐	identify objective clusters for which overall performance was low? [D5]
☐	☐	☐	☐	☐	report normative information for objective clusters? [D6]
					The following items refer to the reporting of subject matter summary reports. Does (should) the report system. . .
☐	☐	☐	☐	☐	report individual student scores for each subject? [E1]
☐	☐	☐	☐	☐	report average scores for each subject? [E2]
☐	☐	☐	☐	☐	report past performance for each subject? [E3]
☐	☐	☐	☐	☐	report normative information for each subject? [E4]
					The following items refer to the provision of specialized services. Does (should) the report system. . .
☐	☐	☐	☐	☐	include a standard statement from the superintendent or other district official? [F1]
☐	☐	☐	☐	☐	include teacher comments about the student? [F2]
☐	☐	☐	☐	☐	highlight data on selected "key" objectives? [F3]
☐	☐	☐	☐	☐	provide individual reports for student record files? [F4]
☐	☐	☐	☐	☐	report performance in relation to aptitude? [F5]
					The following items refer to the provision of the general, standard services. Does (should) the report system. . .
☐	☐	☐	☐	☐	identify the audience for whom the report is intended, the type of information included, and the test data? [G1]
☐	☐	☐	☐	☐	report normative information as a band or as a number with an error term? [G2]
☐	☐	☐	☐	☐	include all information relating to one test or subtest on the same page? [G3]
☐	☐	☐	☐	☐	devote a section of the report to the generalizability of the scores? [G4]
☐	☐	☐	☐	☐	provide a telephone number to call for aid in interpreting the reports? [G5]
☐	☐	☐	☐	☐	provide narrative explanations of data? [G6]

BEST COPY AVAILABLE

SCHOOL DISTRICT SCORE REPORTING SURVEY

Name of Respondent (Optional):_____

Title/Responsibilities:_____

Does your school district have a non norm-referenced (e.g., domain-referenced, objective-referenced, or criterion-referenced) district-wide assessment program?

Check one box: ☐ Yes [A1]

☐ No

DIRECTIONS:

Please take a few minutes to answer the following survey. If you answered "no" to the questions above, please skip to page 2 of the survey and respond according to what you, personally and professionally, believe would make for a good score reporting system for your school district. If you answered "yes" to the question above, please complete the section below before completing the remainder of the survey. Complete the survey based on the score reporting system currently being used by your school district.

Complete the following questions on this page only if you answered "yes" to the question above.

Circle the grade levels assessed (circle as many as apply):

PreK K 1 2 3 4 5 6 7 8 9 10 11 12 [B1-14]

Check the types of reports generated (check as many as apply):

☐ Classroom teacher reports ☐ Building level reports

☐ Individual student (parent) reports ☐ District level reports [B15-19]

☐ Individual subject matter summary reports

Please add any comments below regarding your score reporting system which might be useful in interpreting your responses to this survey.

BEST COPY AVAILABLE

DIRECTIONS

If you answered "yes":

For the rest of the survey, respond to each item in terms of what your reporting system does. Respond for each level of reporting (e.g., teacher report, building report) by placing a check mark ("↘") in the appropriate boxes if this item is part of your score reporting system.

If you answered "no":

Since you do not have a score reporting system, you should place a check mark ("↘") in the appropriate box(es) for each level of reporting (e.g., teacher report, building report) If you feel that item would be desirable in a score reporting system.

Please enclose a copy of any score reporting forms you are currently using. THANK YOU for your participation in this effort. Please return this form and any score reports you choose to send to:

Dr. Joseph F. Haenn
Senior Research Associate
Maryland Score Report Development Project
RMC Research Corporation
108 S. Churton Street
Hillsborough, NC 27278
(919) 732-6121

Student (Parent) Report	Class (Teacher) Report	Building Report	District Report	Items: In order to meet our informational needs, does (should) the score reporting system at each indicated reporting level. . .	
S	C	B	D		
☐	☐	☐	☐	indicate the nuber of items tested per objective?	[C1]
☐	☐	☐	☐	indicate the cut-off score for each objective?	[C2]
☐	☐	☐	☐	identify the responses chosen for incorrectly answered items?	[C3]
☐	☐	☐	☐	report the names of masters and non-masters by objective ?	[C4]
☐	☐	☐	☐	provide diagnostic statements about performance on non-mastered objectives?	[C5]
☐	☐	☐	☐	report number or percent of masters for each objective?	[C6]
☐	☐	☐	☐	identify mastered and non-mastered objectives?	[C7]
☐	☐	☐	☐	report objectives on which performance was low?	[C8]
☐	☐	☐	☐	report average for each objective?	[C9]
☐	☐	☐	☐	report the number of students tested for each objective?	[C10]
☐	☐	☐	☐	report past performances on the same or related objective?	[C11]
☐	☐	☐	☐	report frquency distribution of percent of objectives mastered?	[C12]
☐	☐	☐	☐	report normative information for each objective?	[C13]

Please turn the page.

Student (Parent) Report S	Class (Teacher) Report C	Building Report B	District Report D	The following items refer to the reporting of clusters of objectives or subtests. Does (should) the report system. . .	
☐	☐	☐	☐	report individual scores on each objective cluster?	[D1]
☐	☐	☐	☐	report average score on each objective cluster?	[D2]
☐	☐	☐	☐	report number of percent of objectives mastered for each objective cluster?	[D3]
☐	☐	☐	☐	report past performance on each objective cluster?	[D4]
☐	☐	☐	☐	identify objective clusters for which overall performance was low?	[D5]
☐	☐	☐	☐	report normative information for objective clusters?	[D6]

BEST COPY AVAILABLE

S	C	B	D	The following items refer to the reporting of subject matter summary reports. Does (should) the report system. . .	
☐	☐	☐	☐	report individual student scores for each subject?	[E1]
☐	☐	☐	☐	report average scores for each subject?	[E2]
☐	☐	☐	☐	report past performance for each subject?	[E3]
☐	☐	☐	☐	report normative information for each subject?	[E4]

S	C	B	D	The following items refer to the provision of specialized services. Does (should) the report system. . .	
☐	☐	☐	☐	include a standard statement from the superintendent or other district official?	[F1]
☐	☐	☐	☐	include teacher comments about the student?	[F2]
☐	☐	☐	☐	highlight data on selected "key" objectives?	[F3]
☐	☐	☐	☐	provide individual reports for student record files?	[F4]
☐	☐	☐	☐	report performance in relation to aptitude?	[F5]

S	C	B	D	The following items refer to the provision of the general, standard services. Does (should) the report system. . .	
☐	☐	☐	☐	identify the audience for whom the report is intended, the type of information included, and the test data?	[G1]
☐	☐	☐	☐	report normative information as a band or as a number with an error term?	[G2]
☐	☐	☐	☐	include all information relating to one test or subtest on the same page?	[G3]
☐	☐	☐	☐	devote a section of the report to the generalizability of the scores?	[G4]
☐	☐	☐	☐	provide a telephone number to call for aid in interpreting the reports?	[G5]
☐	☐	☐	☐	provide narrative explanations of data?	[G6]

APPENDIX B

FOLLOW-UP TELEPHONE INTERVIEW FORMS

Follow-up Questions

State_____ Date_____

1. How are results used?

   by State

   by Districts

2. What kinds of support do you provide in report interpretation (e.g., user's manuals, objective-materials match, workshops, other)?

(Follow-up: no. of workshops/yr., nature of manuals, etc.)

3. What changes have you made in system in last few years...why?

4. What kinds of abuses, misuses, or misinterpretations occur in your state?

Follow-Up Questions

District_____ Date_____

1. How are results used?

   by District

   by Schools

   by Teachers

2. What types of questions about reports do you get from:

   Parents

   teachers

   principals

   other district staff

3. How relevant are score reports to local teaching objectives?

4. What are major types of misinterpretation occurring in your district?